DOCUMENT RESUME

ED 312 628 CS 009 839

AUTHOR Birger, Barbara
TITLE A Comparison of Performance on a Silent Informal Reading Inventory and Achievement on a Standardized Reading Test.
PUB DATE Dec 89
NOTE 50p.; M.A. Thesis, Kean College.
PUB TYPE Reports - Research/Technical (143) -- Dissertations/Theses - Master Theses (042) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Comparative Analysis; Grade 4; *Informal Reading Inventories; Intermediate Grades; Reading Achievement; Reading Comprehension; Reading Research; *Reading Tests; *Standardized Tests; *Test Validity
IDENTIFIERS California Achievement Tests

ABSTRACT

A study compared performance on a silent informal reading inventory and achievement on a standardized reading test, focusing on students' individual skill needs and how effective tests were in identifying specific strengths and weaknesses in these skill areas. Subjects, 25 students entering fourth grade in a primarily middle class suburban community, were administered a silent informal reading inventory (IRI) after reading a passage. Subjects' most recent scores on the California Achievement Test (CAT) were compared to the results of the IRI. Results indicated that performance on a silent informal reading inventory do not correlate highly with achievement on a standardized reading test. Findings suggest that assessing student progress in reading ability for appropriate instruction is best achieved by a combination of the normed approach (as in the CAT) and a diagnostic-prescriptive approach (as in the IRI). (One table of data is included; 51 references, the evaluation instrument, data for reading comprehension instructional levels is appended. (RS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED312628

A Comparison of Performance On A Silent Informal Reading Inventory and Achievement On A Standardized Reading Test

By

Barbara Birger

Accepted 12/4/89

Presented in Partial Fulfillment of the Requirements for the Master of Arts Degree - Reading Specialization Program at Kean College of New Jersey

December 1989

CS009839

BEST COPY AVAILABLE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Barbara Birger

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TABLE OF CONTENTS

		Page
I.	Abstract	ii
II.	Acknowledgements	iii
III.	List of Tables	iv
IV.	A Comparison of Performance On A Silent Informal Reading Inventory and Achievement On A Standardized Reading Test	
	Introduction	1
	Hypothesis	5
	Definitions	6
	Procedures	7
	Results	10
	Conclusions	12
V.	A Comparison of Performance On A Silent Informal Reading Inventory and Achievement On A Standardized Reading Test Related Research	14
VI.	References	35
VII.	Appendix	39

ABSTRACT

The purpose of this study was to make a comparision of performance on a silent informal reading inventory and achievement on a standardized reading test. The focus was on the individual skill needs of the tested students and how effective the tests were in terms of identifying specific strengths and weaknesses in these skill areas. An attempt to determine if students performed differently on these two tests in relation to reading achievement was made.

The results indicated that there was a significant difference in the test results on an informal reading inventory and a standardized reading test.

ACKNOWLEDGEMENTS

I would like to express my appreciation to my husband, John, and my children, Marc and Daryl, for their support and understanding while I was writing this thesis.

I would like to thank Dr. Putnam for her guidance and assistance in Clinical Internship for instruction on the proper administration and analysis of informal reading inventories.

A special thank you to Dr. Mazurkiewicz for his guidance and patience throughout this project.

Special thanks must also be given to the students who cooperated and assisted me in the completion of this paper.

LIST OF TABLES

		Page
I.	Mean, Standard Deviation, and r of the instructional levels scores of the Burns and Roe Informal Reading Inventory and the California Achievement Test of the Specific Sample	10

The diagnosis of reading ability is best achieved using standardized tests and informal procedures. Classroom teachers constantly make informal diagnosis noting success and errors in word recognition, ability to anticipate words using meaning clues, and tendency to regress when unsure of meaning or to correct word recognition. (Bond and Tinker 1984)

Reading difficulty is viewed as a gap between cognitive ability and reading performance. It depends on the interaction among the specific reader, the text, and instructional factors. (Wixson 1987) Reading diagnosis by classroom teachers focus on the next steps for improving the reading of the individual for overcoming the difficulty. (Chall 1987)

Standardized tests compare a student's performance with the performance of their classmates. This determines placement in Basic Skills instruction classes by using state percentile cut offs. It also qualifies students for the President's Academic Fitness Awards when combined with class achievement. For an individual student, the use of NCEs allows meaningful comparisons between different achievement test batteries and between different tests within the same test battery. The NCEs obtained by different groups of students on the same test or test

battery may also be averaged for the purpose of comparison.

On the other hand, informal reading inventories determine how each student can read without comparison to others. It can diagnose specific strengths and weaknesses.

The informal reading inventory (IRI) is administered individually to determine a child's oral and silent reading abilities. The student reads orally and silently from increasingly difficult content until the material becomes frustrating, either in terms of accuracy of pronunciation or understanding of ideas in context. It reveals specific areas of difficulty in reading, (Zintz 1977) It shows readers do not have one single reading level, but they have three reading levels, each appropriate for reading of different kinds of text for different purposes.

In determining the independent level, the student can read text easily, without help. Comprehension is excellent, and silent reading is rapid because almost all of the vocabulary is recognized and understood at sight. Oral reading is fluent and errors rarely interfere with comprehension. This level best meets a student's needs for reading for pleasure, information, and enrichment as well as homework, seatwork, and tests to be completed alone.

In determining the instructional level we find the

material is not really easy but it is still comfortable. The student is challenged and will benefit from instruction. Comprehension is good, but help is needed as readiness to understand some concepts. Silent reading rate is fairly rapid, although usually slower than at the independent level. Some word analysis is needed, but most words are recognized at sight. Oral reading is smooth and accurate with context helping comprehension. This level is where direct instruction will take place. These materials can be completed with teacher help and guidance.

The frustration level is where the material is too difficult in vocabulary or concepts to be read successfully. Comprehension is poor, with ideas forgotten or misunderstood. Both oral and silent readings are slow and labored, with word recognition poor. This level is to be avoided in instruction.

Since IRIs are individually administered the teacher is able to observe students directly while they are responding to reading tasks. Best performance is in a relaxed environment where a good rapport encourages a cooperative attitude. The teacher's rapport and a relatively uninterrupted environment with no disturbances outside the classroom are necessary. Individual sessions with students reduce the competition that frustrates the

low achieving readers. (Wilson & Cleland 1985) Accurate results on an IRI depend on the teacher administering the test. It involves subjectivity. The type of comprehension questions used affects the results.

Comprehension questions measure factual information, inferences, main idea, sequence, and vocabulary. Oral and silent passages could be a continuous reading of the same story.

The CAT is a group standardized test and measures abilities involved in silent reading. It assesses vocabulary and comprehension in separate subskills. Reading vocabulary items measure categorization, same and opposite meaning words, and word affixes. Reading comprehension items measure skills in understanding sentence meaning, passage details, character analysis, main ideas, generalizations, written forms, and author techniques.

Results from standardized testing may show two children scoring at the same level in reading and yet have different sets of skills contributing to that result. Because of this diversity of skills, additional testing may be desirable to locate strengths and weaknesses. (Potter and Rae 1973)

Diagnosing a child's past learning and present

capabilities is a direct attempt to determine effective teaching strategies. It will be the basis of curriculum decisions with regard to materials, appropriate techniques in the pacing of the lesson, reassurance needed, or the amount of encouragement needed to do their best. Some children learn a substantial amount of material incidentally while other children need carefully structured experiences to acquire the same skills.

Effective reading instruction is impossible without diagnosis on a formal and informal level. Knowing strengths and weaknesses help meet the student's individual needs and therefore would create less frustrating reading experiences for the child. (Mazurkiewicz 1968)

Yet research suggests that both informal and standardized test data should be provided to the teacher to improve the match between ability and instruction. However correlational data are not uniform and range from inadequate to high correlations.

Hypothesis

To provide additional evidence concerning this topic, a study was conducted on the hypothesis that the correlation of performance on a silent informal reading inventory with achievement on a standardized reading test will be high.

Definitions

IRI: An Informal Reading Inventory is a formally published test designed to help discover the levels of reading material that pupils can read with and without teacher assistance, the reading levels at which pupils should not be able to function based on their ability to comprehend material that is read to them.

INDEPENDENT READING LEVEL: The highest level at which a child can read fluently and with personal satisfaction without help.

INSTRUCTIONAL READING LEVEL: The highest level at which a child can do independent work with the usual readiness from the teacher.

FRUSTRATION LEVEL: The lowest level at which obvious difficulties cause confusion, frustration, and tension in the reading situation.

SILENT "AT SIGHT" READING: Reading without previous preparation of a teacher selected story and a short objective test covering the material.

CAT: California Achievement Test, a series of test batteries that combine the most important and useful characteristics of a norm-referenced and criterion referenced test, providing ranking and instructional needs of students

PERCENTILES: A range from 1-99, indicating how well a student has tested in relation to other students in the same grade. A percentile rank indicates the percentage of scores in a norm group that fall below a given student's scale score.

NCE: Normal Curve Equivalent - an equal interval scale that, unlike percentages, would permit averaging and comparison of test data. They range from 1 - 99.

GRADE EQUIVALENTS: Scale that ranges from .0 to 12.9, representing the thirteen years of school and the ten months in the traditional school year. It represents the grade and month in school of students in the norm group whose test performance is theoretically equivalent to the test performance of a given student.

Methods and Procedures

The subjects for comparing these two types of tests were twenty-five students entering fourth grade in a primarily middle class, suburban community. The students were heterogeneously grouped and equally divided into one of two fourth grade classes by academic achievement, race, and sex.

Each student was initially administered a word recognition test in order to locate the grade level at which to begin the inventory. A word list was used that

was at least two years below the student's present grade level. If any words were misread the student was directed to drop to an easier level until no errors were made.

The list was also used to determine probable independent, instructional, and frustration levels. The level at which a student missed no more than two out of twenty words was considered the independent reading level; four errors on a list indicated the probable instructional level; and five or more errors identified the level at which reading material was too difficult.

The highest-level list on which the pupil made no errors indicated the level at which the examiner then initiated the IRI graded passage. This technique was employed to guard against frustrating the students with material that might be too difficult.

The teacher determined independent, instructional, and frustrational levels for each student using the Burns and Roe silent informal reading inventory section. But only the instructional and frustration levels were used in statistical analysis.

The pupils were instructed to read each passage only once but carefully because they would be asked to answer comprehension questions. They were never given assistance

and passages were never available during questioning. Introductory motivational statements were made before reading each passage. Question probes were used in assessing children's responses to comprehension questions. When an answer to a comprehension question was considered incomplete, the child was asked to "tell more" about the content being questioned or how the child arrived at the answer. The responses yield information about the child's thinking process in answering the questions.

Criteria for scoring was in the form of a percentage based on the number of errors made in answering comprehension questions. Independent levels were reached at 90 percent or higher, instructional levels at 75 percent or higher, and frustration below 50 percent.

The results of the CAT were the most current assessment on a standardized level. These tests were administered by the two previous teachers, the end of the third grade year. Any new student entering fourth grade from outside the school system took the CAT administered by the Basic Skills instructor within the first or second week of school.

The performance on the IRI was correlated to class and individual list reports on the CAT, noting grade equivalence in the total reading results, identified as the frustration level and correlated to its instructional level by use

of a constant (-1.5).

The testing procedure continued approximately five weeks, testing one student per day. Scores were compiled and correlated to determine the relationship between the two instruments.

Results

Table 1 illustrates the data collected (see Appendix B) when comparing the results of performance from an instructional level on the Burns and Roe Informal Reading Inventory and the calculated instructional level on the Reading Total from the California Achievement Test.

Table 1

Mean, Standard Deviation, and r of the IRI and the CAT test results

Sample	Mean	SD	r
IRI	4.46	1.77	0.62
CAT	4.26	2.63	

To analyze the results, mean scores were compiled and correlated to determine if there was any difference between the results of the two tests or whether they measured essentially the same thing.

The data collected indicated that although both tests had a similar mean, there was a significant standard deviation due to a wide academic range in ability.

The Pearson Product Moment Formula was applied to determine the correlation between the IRI scores and the CAT scores. As we saw in Table 1, the sample showed a mean difference of .20 between the two measures and a standard deviation mean difference of -.86.

The correlation co-efficient of 0.62 between the IRI and the CAT was obtained, indicating that only 38% of the time are the IRI and the CAT measuring the same thing.

Although the mean IRI test score was 4.46, which is on grade level, the individual scores ranged from a grade equivalent of .5 to a grade equivalent of 7.5. Also noteworthy is the range of grade equivalence on the CAT. The mean was 4.2, however, the range was from a low of 1.5 to a high of 11.1.

Conclusions

The results of this study do not support the hypothesis that performance on a silent informal reading inventory correlates highly with achievement on a standardized reading test. An examination of this data indicates that the two tests are measuring reading differently and that only 38% of the time are the two tests providing evidence on the same thing. Therefore it is concluded, that assessing student progress in reading ability for appropriate instruction is best achieved by a combination of the normed approach as in the CAT where student progress is evaluated in terms of ability of other students and a diagnostic-prescriptive approach as in the IRI in terms of specific skill mastery on an individual level.

Differences may have resulted from the fact that standardized reading tests such as the CAT are administered in a group situation where as the IRI is administered in a one-to-one setting.

The standardized test scores were based upon vocabulary (word meaning) and passage comprehension performance with relation to grade equivalent scores, minus the constant 1.5, to determine an instructional level. IRI scores were based upon word recognition and passage comprehension of a graded silent reading selection to determine an instruction level.

Standardized tests and IRI results depend on the contents of the paragraphs in the passage comprehension. Processing of information may depend upon interest level, background information, and comprehension level.

IRI test results might increase if the method of administration encouraged readers to reinspect the selection before answering. There could be a significant change from frustration to instructional levels as this would be a true measure of reading skills rather than memory factors. This process is possible when taking the CAT but sometimes due to the fact that this is a timed test, students who read at a low rate of speed may not have time to be able to benefit from this opportunity as they will not complete the test in the required time. Since an IRI is not done in a timed environment, administration is more relaxed and students could skim or scan if allowed to reinspect for specific answers that are not in their immediate memory.

As previously mentioned, these CAT results can show two children scoring at the same level in reading by grade equivalence yet have different sets of skills contributing to that result. Due to this diversity of skills, additional testing should be used to diagnose strengths and weaknesses.

A Comparison Of Performance On A Silent Informal
Reading Inventory and Achievement On A
Standardized Reading Test
Related Research

Assessment in reading should be to study, evaluate, or diagnose reading behavior, to monitor student progress, to supplement and confirm information gained from standardized tests, or to obtain information not readily available from other sources.

Mc Cormick (1987) suggests that IRIs can be used as both a survey test and a diagnostic test. Research shows that grade scores obtained on an IRI are closer to the actual classroom performance of students than scores from standardized survey tests. Alexander and Heathington (1988) found survey tests more meaningful for group comparisons than for diagnosing individuals, since they do not yield enough information on specific skills to permit an adequate diagnosis. (Mercer and Mercer 1985) (Ekwall and Shanker 1988)

Bader and Wiesendanger (1989) consider IRIs to be an excellent tool to evaluate reading behavior in depth. Due to the informal appraisal the student is confident and willing to risk error, able to make substitutions, tries to be fluent, and provides observable behavior. Trial teaching can occur on the spot and the reader's responses can be evaluated. When an untimed word recognition test is administered the missed words can be presented for decoding. This provides information on the student's

decoding skills without time pressure. It also gives the teacher an idea of which word recognition techniques the child can already use in decoding words. They obtained a .93 correlation between reading levels obtained with a published IRI and classroom teachers' placements.

Assessments should be as closely related to instructional procedures as possible. Norm-referenced tests are not reality oriented. Comprehension on this kind of test is determined by multiple choice questions on short passages in a timed situation. Day-to-day classroom instruction comprehension is generally checked by recall questions following the reading of longer passages in untimed situations. (Alexander and Heathington 1988)

Readence and Martin (1988) are also concerned that standardized reading tests do not represent real reading tasks and situations like reading tasks students are assigned in the classroom. After reading short paragraphs the multiple choice question format limits a complete picture of students' ability to comprehend. It also allows students to get some items correct by guessing. Scores are thus inflated by adding the correct items from reading comprehension to the guessed items. With four-alternative responses, a student's guesses on all items could lead to a 25 percent rate of correct response.

(Gillet and Temple 1986)

The difference in performance on comprehension questions using a multiple choice format and a probed recall format varies by grade level, but multiple choice questions are always easier. Multiple choice question scores give data for pupils' instructional reading level, whereas probed recall questions scores appear to represent pupils' frustration reading levels. (Blanchard, et al. 1983)

Readence and Martin (1988) recommend that if teachers use an informal reading inventory to gain the most relevant diagnostic information, the inventory should be designed to use passages from the textbooks students use rather than published informal reading inventories. They feel these IRIs are "contrived text" and not what the students use every day in reading assignments. Masztal and Smith (1984) researched the use of IRIs by elementary teachers. The respondents indicated that they do utilize the IRI in their classroom, they vary in the kind of IRI used, but the one most often administered seemed to be those that accompany the basal series.

Gillet and Temple (1986) have identified shortcomings of commercially prepared IRIs and basal reading series IRIs. Basal IRIs contain passages from the reading series

and students may have read a story previously thus the resulting scores would be falsely inflated. Commercial IRIs conform to readability levels or come from other basal readers, but the quality varies. Some passages are too short and limit the ideas and number of questions.

Jongsma and Jongsma (1981) recommends that teachers should weigh the advantages and limitations of commercial tests against teacher developed inventories or basal inventories. They feel the teacher-developed inventories tend to be more closely related to the instructional materials and practices being used in the classroom.

IRIs can be developed and administered for science and social studies content. It can tell the teacher whether the student can read the text independently. Many texts at the intermediate level have a wide range with readability levels above grade expectations. The teacher can not always change the content but can alter techniques to cover it. (Mercer and Mercer 1985)

The Burns and Roe (1985) Reading Inventory word list items were randomly selected from two basal reading series (Rand Mc Nally and Scott Foresman). The lists were administered to students on, above, and below the levels intended. The percent of students who successfully pronounced each word was calculated for each of the

levels. Words that failed to meet an 80% criteria were replaced on any level.

The comprehension reading passages were checked for readability level using the Spache Readability Formula, pre-primer through grade three, and the Fry Readability Graph for grades four through twelve. This included fiction, non-fiction, and only excerpts that made sense out of the context of the entire story . The motivational statement preceding each passage included any necessary background information.

Comprehension questions cover main idea, detail, sequence, cause and effect, inference, and vocabulary. There are two types of main idea questions, one that asks for the main idea as a main idea or a title. The other states a question that must be answered with the main idea.

Classification of questions can be difficult, but the way a question is answered, not the way the question is asked, determines a detail or inference question. All questions that are not directly answered by the passage are inference questions.

In the second edition of this informal test passages were replaced to include current concerns, represent minority groups, and raise the overall interest level of

passages.

A field testing revealed the difficulty level of the passages did increase in each of the four forms. The levels found for the students who were tested agreed closely with their teachers' judgments of the levels at which they were performing.

The examiner must make an accurate record of the exact way in which the child reads the oral selection noting miscues. One should distinguish between trivial and significant errors. By erring a student reveals something of word recognition and comprehension strategies being used. (Miller and Mc Kenna, 1989) It should be noted if students fail to use semantic or meaning clues. This might signal a need for vocabulary and concept development and use of context clues. If a student fails to use syntactic or grammatical information the student's spoken dialect may be unlike standard English. Disregard of graphic or symbol clues would indicate a need for phonics and structural analysis instruction. Miscues would be mispronunciations, substitutions, refusal to pronounce, insertions, omissions, repetitions, or reversals.

In addition to determining word recognition, oral reading miscues, and answers to comprehension questions, evaluation of performance in fluency is equally important.

In an IRI the fluency is the smoothness with which both oral and silent reading proceed. Judgments about oral fluency can be made using intonational factors, but rate and comprehension must be used to make judgments about the fluency with which silent reading discloses. (Guszak 1970) Better readers vary their reading rates according to material and purpose. Timing a child during oral and/or silent reading could be used for direct comparison or measure against the typical speed for children their age or grade level. Children who read accurately but at a slow rate would benefit placement in a lower group or reading level. (Alexander and Heathington (1988)

Another kind of data obtained from IRIs should be the observation of related tension signs and behavioral characteristics. When observing the silent reading selection careful observation is essential if the child's reading performance is to be fully interpreted. Silent reading behavior should be noted if the student relies on use of picture clues, shows lip or head movement, finger pointing, or vocalization. (Johnson, M. et al. 1987) (Ekwall and Shanker 1988) (Alexander and Heathington 1988) Especially at the frustration level one could observe voice tone change, arhythmical or word-by-word reading, lack of expression in oral reading, inaccurate

observation of punctuation, frequent requests for examiner help, noninterest in the selection, yawning, fatigue, or refusal to continue.

In determining reading levels a teacher should determine the independent level when the student scores high in both word recognition and comprehension as well as being free of tension signs. When fewer than three signs of anxiety are observed it would benefit the student to be placed on the instructional level due to the observable need for help. On the instructional level the student should display two or fewer signs of anxiety. The frustrational level would be four or more observable signs.

In examining the progress and reading attitudes of children placed in frustration level materials it was noted that student achievement in reading is positively affected by placement in materials that produce low error rates between two and five percent. Readers placed in reading material which produce error rates greater than five percent tend to be more frequently off-task during instruction. (Allington, 1984)

If we compare a present reading achievement level based on the results of the administered standardized test with the instructional level obtained from the IRI we can feel confident in placing a student on that level. But, if

we see a great discrepancy between the results of two tests one should carefully review the administration of the tests that may have resulted in a score not accurately reflecting the student's reading level. (Mc Cormick 1987)

Brecht (1977) suggests that achievement levels based upon errors made during oral reading from sight will not be representative of actual reading ability. To get the best measure of a child's instructional reading level it is best to allow them to read the selection silently first. More than two-thirds of students tested scored significantly higher when the oral reading took place after an initial silent reading. This reduced errors to the extent that they attained oral reading scores at least one level above performance on material not preread.

Mc Kenna (1983) also agrees about oral reading at sight because it is contrary to the usual classroom practice of asking children to read aloud only after they have read the material silently.

Kender and Rubenstein (1977) feel that reinspection of a passage by a reader should be part of the administration of an informal reading inventory. This is more closely linked to reading performance on standardized tests or individualized assignments. A reader keeps the material and is instructed to feel free to refer to it when

answering questions. Determination of the reader's functional level is therefore not contingent upon just recall but also understanding. Taking away the material when questioned, readers may comprehend but fail to recall the information, resulting in our underestimating their comprehension. (Cardarelli 1988) Otherwise an IRI would be more of a test of memory than of reading comprehension.

This produces greater demands on long term memory rather than search strategies and logical reasoning skills. In various studies over half of the students had made gains sufficient to change their functional reading level from frustration to instructional. Gonzales and Elijah (1975) suggest that there is a reduction in the number of errors when passages are reread. Even though the number of errors do change the pattern of errors remains constant. When material is too difficult, reinspection is not going to make it understandable as identified in 30% of the students.

In a standardized reading comprehension test multiple choice questions represent a recognition task with memory search and retrieval processes based on written clues. Poor comprehension could be because of poorly worded questions and interference from the many written

clues available. (Blanchard, et al. 1983)

In an IRI comprehension questions also involve memory search and retrieval, but based only on verbal cues from oral and silent reading. Using probed recall there are no responses to choose from. A student can comprehend the passage but score poorly because of poor recall, a poorly worded question, inability to form a verbal response, inability to articulate an acceptable response, or the inability of the teacher to evaluate a correct response.

Research shows that multiple choice questions produced statistically significantly fewer errors than did probed recall questions.

Recently many IRIs are commercially available and teacher educators instruct their college students to utelize the IRI for diagnosis in special reading classes, clinics, or classrooms. (Masztal and Smith 1984)

Since the IRI must be administered individually they are more time-consuming than group-administered standardized tests. Many teachers feel using IRIs took too long to administer, but reading and learning disability specialists using them in a clinical situation had no problems dealing with administration time. (Searls 1988)

Harris and Lalik (1987) surveyed teachers from Virginia and found that building and district level policies may

supersede the decision of individual teachers. They are thwarted in their attempts to implement techniques taught in teacher education programs. Classrooms were either self-contained and did not need IRI use, reading group placement was done by the reading specialist or the principal, or was continued from the previous year. Successful completion of a given level the previous year may not represent mastery of that level.

Schell and Hanna (1981) feel IRIs fail to provide accurate, reliable, comparable scores on subskills of reading. There is a disagreement between authors of IRIs and teachers on classification of comprehension questions. Using five classification categories they were unable to conclude that a question necessarily measures the comprehension category that the author says it does. Therefore one could not determine strengths and weaknesses on the basis of the author's labels.

Teachers need help to become experts at evaluating the process of literacy development. The bulk of instructional decision making takes place in the classroom on a moment to moment basis. Teachers must respond to the students' needs they have evaluated. They must have the ability to recognize patterns. They must know what patterns to look for and the conditions under which the patterns

are likely to occur. (Johnston 1987)

Comprehension skill patterns can be noted if problems in answering a particular type of question are seen at different grade levels. This would allow for the design of appropriate comprehension activities for students according to their individual needs. (Gillet and Temple 1986) Teachers must use critical judgment when they use IRI questions if the subskill groups are not clear, and one must consider whether the questions are appropriately labeled by type.

Due to use of raw or percent scores the IRI should contain data showing item difficulty values across categories according to one or more measures of readability. (Gerke 1980)

Interest level significantly affects comprehesion scores. Low interest in a given passage can cause underestimation of a student's ability due to a deflated score. Students pay closer attention to materials that interest them. This leads them to develop more background, ensuring more insightful reading. Ideal interest areas are those that are moderate to high for both sexes at the grade level for which the IRI is intended. (Mc Kenna 1983)

Some questions are termed passage independent and can be answered without having read the passage. This would

then appear to be a strength when in reality it was answered using prior knowledge or use of clues within the question.

Most comprehension questions seem to function as an oral exam rather than to develop understanding. It should be a time to use quality reasoning. Recall of details or appropriate interpretation could take place without understanding context. (Nessel 1987) Even achievement tests tend to call upon factual material rather than higher-level thinking skills. These tests do not measure learning rate, interest, or creative thinking. Students do not have the opportunity to show important comprehending behaviors such as justifying answer choice, posing their own questions, visualizing situations, or prediction. (Readence and Moore 1983)

The California Achievement Test is a survey test to determine an approximate reading level and identify those with severe problems. It determines overall levels of proficiency rather than pinpointing specific skill strengths and needs. A multi-reference report will identify the item analysis of errors and the instructor would have to identify patterns of skill errors. These reports would have to be requested for each individual whereas most often these reports would be issued only for

those qualifying for basic skills instruction.

Survey standardized reading tests tend to be better estimates of true ability for students with average performance than for the very good or very poor readers. These scores are less accurate. They indicate that a reading difficulty exists, but does not reveal the nature or degree of the difficulty. (Gillet and Temple 1986)

Weiner (1979) studied five different standardized reading achievement tests to assess the interchangeability of the tests when grade-equivalent cutoff scores were used as pass-fail criteria. He found differences in passing rates among the tests which showed the tests to be nonequivalent in their outcomes. Recommendations were made to conduct only one test. Local validation studies should be conducted to identify the level of achievement appropriate for use as a cutoff score in the given system. Use of the test publisher's grade-equivalent scores may not be appropriate in each instance. Different norm groups in original validation of the tests supply restricted grade-equivalent scores.

Grade scores overestimate reading ability, representing frustration levels rather than the supposed instruction level. (Gillet and Temple 1986) These scores should not be used to place children in textbook levels. A

common practice is to subtract one year from the score to determine the student's instructional level based on a standardized test score, and an additional year can be subtracted to find the independent level. Many argue that this is statistically incorrect.

Mc Cormick (1987) noted the standardized test score could be an underestimate for those students whose reading ability is adequate but whose test-taking skills are poor. Converting scores could deny eligibility for students to enter programs they need or even place some students in classes they don't need.

Crowell (1983) designed a study to identify types of questions in six major standardized tests including the CAT. In a "text explicit relationship", both the question and the answer can be derived from the text and is right there. In a "text implicit relationship", both the question and answer can be derived from the text, but one or more steps of inference are required to get from the question to the answer therefore one must think and search. In a "script implicit relationship", the answer cannot be derived from the text but can be determined with reference to background knowledge in his or her own memory.

The results indicated that the California Achievement Test attended to just two kinds of question-answer

relatonships, text explicit and text implicit. Script implicit items were few in number. The percentage of text implicit questions increased gradually in the upper grades. In comparison to the other five standardized tests the CAT ranked fourth comparing percentages of higher level questions.

Parents feel standardized tests are important but they misunderstand the meaning of the test scores. They prefer report styles that indicate how the child had improved over past performance. (Dreher and Singer 1985) Alexander and Heathington (1988) suggest that grade-equivalent scores should not be reported to parents since they are too often misunderstood to represent a precise, measurable level of achievement.

Parents and school staff would prefer to see the most difficult item a child could successfully complete and a comparison of the child to his or her own past performance rather than only to the norm group. This would be more informative to parents and more useful in instruction for teachers.

Tests used to measure reading achievement do not reflect recent advances in our understanding of the reading process. If we look at reading as prior knowledge, comprehension, metacognition habits, and

attitudes, we see the need for new assessment techniques based on these attributes. Our present tests check comprehension by using lots of short passages on lots of topics. They rely on too many literal comprehension test items, and multiple choice items with only one correct answer. They rarely go beyond finding the main idea, and rarely consider fluency as an index of skilled reading. Reading comprehension should use prior knowledge, inference, a complete story structure, fluency, with application and use of knowledge in new situations. (Valencia and Pearson 1987)

IRIs were originally created for the purpose of diagnosis and placement based directly on the materials used for instruction responsive to the needs of the students in reaction to the recognized limitations in group standardized assessment procedures. (Barr and Sadow 1985) These devices should be more sensitive to classroom instruction and would be more accurate in assessing students' strengths and weaknesses and their instructional levels. (Fuchs, et al., 1982)

Assessment should be made of both weaknesses and strengths. Students should know what they do well and should be encouraged to build on their strenghs providing the motivation while they work to overcome weaknesses.

Then weaknesses can be analyzed to diagnose the reasons behind these abilities. The will or desire to learn could be affected by attitudes, motivation, interests, self-concept, or emotional/personality maladjustment. The process of reasoning or recognizing could be affected by problems with attention, perception, conceptualization, intelligence, and metacognition. Determining the deficiencies will make correction easier. (Alexander and Heathington 1988)

Despite the use of Informal Reading Inventories or similar techniques dating back to the 1920's, IRIs have been highly criticized. This has centered upon scoring criteria, validity of reading levels, scaling of passages, definition of errors for oral reading, accuracy and consistency of teacher judgments, and validity and reliability. These problems confronting users are common to other approaches to reading assessment as well and therefore we should not abandon the use of IRIs. (Pikulski and Shanahan 1982)

Helgren-Lempesis and Mangrum (1988) advise further research to determine the reliability to examine how closely the results from different IRIs approximate each other and the reliability of alternate form use. We also need to know how accurately published IRIs predict reading levels

identified by teachers using IRIs provided by a basal reader or teacher prepared.

Many researchers suggest when using an IRI that one should remember the results are just a sampling behavior. On another set of passages, given on another day, you might get different results. Day to day observations must be considered. Performance may not even be consistent between two basal reader series, since basals often differ in difficulty level and type content.

Readence and Moore (1983) state comprehension is an internal, unobservable behavior. We can only measure behaviors that reflect it. The more methods we use, the better chance we have of making valid assessments. (Ekwall and Shanker 1988)

The reading disabilities and reading level of individual students cannot be accurately assessed using most group tests. Therefore teachers should learn to administer, score, and interpret informal reading inventories.

Readers do not have one single reading level, but they have three reading levels, each appropriate for reading of different kinds of text for different purposes. Teachers continue to gather information both quantitatively and qualitatively relating to the reading ability of students.

Quantitative information is expressed in grade equivalent scores to indicate independent, instructional, and frustration reading levels. Qualitative information concerns the reader's word recognition and comprehension difficulties.

This instrument can identify specific strengths and weaknesses as compared to norm-referenced tests where we see limitations in the conclusions. But as previously noted in research studies, IRI assessment also has numerous difficulties with results.

Hence neither standardized tests nor IRIs should be used exclusively. The tests should be evaluated before making important instructional decisions on the basis of test scores. Direct observation along with test results should help the teacher to plan needed corrective instruction and to provide the student with reading materials suited to their abilities.

References

Alexander, J. Estill and Betty Heathington. Assessing and Correcting Classroom Reading Problems. Illinois:Scott, Foresman and Co., 1988.

Allington, Richard. "Oral Reading," in P. David Pearson (ed.) Handbook of Reading Research. New York: Longman Inc., 1984.

Bader, Lois and Katherine Wiesendanger. "Realizing the Potential of Informal Reading Inventories," Journal of Reading. 32:402-408. February 1989.

Barr, Rebecca and Marilyn Sadow. Reading Diagnosis For Teachers. New York: Longman Inc., 1985.

Blanchard, Jay, et al. "Determining Instructional Reading Level: Standardized Multiple Choice Versus IRI Probed Recall Questions," Journal of Reading. 26:684-689. May 1983.

Bond, Guy and Miles Tinker, et al. Reading Difficulties Their Diagnosis and Correction. New Jersey: Prentice-Hall, 1984.

Brecht, Richard. "Testing Format and Instructional Level With The Informal Reading Inventory," The Reading Teacher. 31:57-59. Oct. 1977.

Burns, Paul, and Betty Roe. Burns/Roe Informal Reading Inventory. Boston: Houghton Mifflin Co., 1985.

California Achievement Test. California: CTB/Mc Graw-Hill, 1985.

Cardarelli, Aldo. "The Influence of Reinspections on Students' IRI Results," The Reading Teacher. 41:664-667. March 1988.

Chall, Jeanne and Mary Curtis. "What Clinical Diagnosis Tells Us About Children's Reading," The Reading Teacher. 40:784-788. April 1987.

Crowell, Doris, et al. "Comprehension Questions: Differences Among Standardized Tests," Journal of Reading. 26:314-319. May 1983.

Dreher, Mariam and Harry Singer. "Parent's Attitudes Toward Reports of Standardized Reading Test Results," The Reading Teacher. 38:624-632. March 1985.

Ekwall, Eldon and James Shanker. Diagnosis and Remediation of the Disabled Reader. Boston:Allyn and Bacon, Inc., 1988.

Gerke, Ray. "Critique of Informal Reading Inventories: Can A Valid Instructional Level Be Obtained," Journal of Reading Behavior. 12:155-158. Summer 1980.

Gillet, Jean and Charles Temple. Understanding Reading Problems - Assessment and Instruction. Boston: Little, Brown and Co., 1986.

Gonzales, Phillip and David Elijah, Jr. "Rereading: Effect on Error Patterns and Performance Levels On The IRI," The Reading Teacher. 28:647-652. April 1975.

Guszak, Frank. "Dilemmas in Informal Reading Assessments," Elementary English. 47:666-670. 1970.

Harris, Larry and Rosary Lalik. "Teachers' Use of Informal Reading Inventories: An Example of School Constraints," The Reading Teacher. 40:624-630. April 1987.

Helgren-Lempesis, Valerie and Charles Mangrum II. "An Analysis of alternate-Form Reliability of Three Commercially-Prepared Informal Reading Inventories," Reading Research Quarterly. 21:209-215. Spring 1986.

Johnson, Marjorie, Roy Kress, and John Pikulski. Informal Reading Inventories. Newark, Delaware: International Reading Associaation, 1987.

Johnston, Peter. "Teachers as Evaluation Experts," The Reading Teacher. 40:744-748. April 1987.

Jongsma, Kathleen and Eugene Jongsma. "Test Review: Commercial Informal Reading Inventories," The Reading Teacher. 34:679-705. March 1981.

Kender, Joseph and Herbert Rubenstein. "Recall Versus Reinspection in IRI Comprehension Tests," The Reading Teacher. 30:776-779. April 1977.

Masztal, Nancy and Lawrence Smith. "Do Teachers Really Administer IRIs?" Reading World. 24:80-83. October 1984.

Mazurkiewicz, Albert, ed. New Perspectives In Reading Instruction. New York: Pitman Pub. Co., 1968.

Mc Cormick, Sandra. Remedial and Clinical Reading Instruction. Ohio: Merrill Pub., 1987.

Mc Kenna, Michael. "Informal Reading Inventories: A Review of the Issues," The Reading Teacher. 36:670-670. March 1983.

Mercer, Cecil and Ann Mercer. Teaching Students With Learning Problems. Ohio:Charles E. Merrill Pub. Co., 1985.

Miller, John and Michael Mc Kenna. Teaching Reading in the Elementary Classroom. Arizona:Gorsuch Scarisbrick, Pub., 1989.

Nessel, Denise. "Reading Comprehension: Asking The Right Questions," Phi Delta Kappan. 68:442-444. February 1987.

Pikulski, John and Timothy Shanahan. Approaches to the Informal Evaluation of Reading. Newark, Delaware: International Reading Association, 1982.

Potter, Thomas and Gwenneth Rae. Informal Reading Diagnosis. New Jersey:Prentice-Hall, 1973.

Readence, John and Michael Martin. "Comprehension Assessment: Alternatives to Standardized Tests," in Susan Mandel Glazer, et al. (eds.) Reexamining Reading Diagnosis - New Trends and Procedures. Delaware: International Reading Assoc., 1988.

Readence, John and David Moore. "Why Questions? A Historical Persepective On Standardized Reading Comprehension Tests," Journal of Reading. 26:306-313. May 1983.

Schell, Leo and Gerald Hanna. "Can IRI's Reveal Strengths and Weaknesses in Comprehension Subskills?" The Reading Teacher. 35:263-268. December 1981.

Searls, Evelyn. "What's The Value of an IRI? Is It Being Used?" Reading Horizons. 29:92-99. Winter, 1988.

Valencia, Sheila and P. David Pearson. "Reading Assessment: Time For A Change," The Reading Teacher. 40:726-732. April 1987.

Weiner, Max and Mara Zibrin. "Dissimilarities in Grade-Equivalent Scores on Different Standardized Tests of Achievement: A Threat to Criterion-Related Validity," Educational and Psychological Measurement. 39:923-928. Winter 1979.

Wilson, Robert and Craig Cleland. Diagnostic and Remedial Reading. Ohio:Bell & Howell Co., 1985.

Wixson, Karen and Charles Peters, et al. "New Directions In Statewide Reading Assessment," The Reading Teacher. 40:749-754. April 1987.

Zintz, Miles V. Corrective Reading. Iowa: William C. Brown Co. Pub., 1977.

Appendices

Form 2–4

SUMMARY — TEACHER

NAME _____ GRADE _____ ORAL _____ SILENT _____ FORM _____

ADMINISTRATOR _____ DATE _____ RATE _____ HIGH _____ AVERAGE _____ LOW _____

INDEPENDENT LEVEL _____ INSTRUCTIONAL LEVEL _____ FRUSTRATION LEVEL _____ CAPACITY LEVEL _____

Strengths and weaknesses in word recognition: include information about use of or failure to use syntactic and graphic clues and about dialect-related miscues, which should not be considered serious recognition problems. Check items on chart. (Use back of form for additional space.)

Single consonants	___	Word middles	___
Consonant blends	___	Word endings	___
Single vowels	___	Prefixes	___
Vowel digraphs	___	Suffixes	___
Consonant digraphs	___	Inflectional endings	___
Diphthongs	___	Compound words	___
Word beginnings	___	Syllabication	___
		Accent	___

Strengths and weaknesses in comprehension: include information about use of or failure to use semantic clues. (Use back of form for additional space.)

Appendix A

WORD RECOGNITION MISCUE ANALYSIS CHART

Miscue	PP	P	1	2	3	4	5	6	7	8	9	10	11	12	Total
Mispronunciation															
Substitution															
Refusal to pronounce															
Insertion															
Omission															
Repetition															
Reversal															
Total															

COMPREHENSION SKILL ANALYSIS CHART

Skill	Number of Questions	Number of Errors	Percent of Errors
Main idea			
Detail			
Sequence			
Cause and effect			
Inference			
Vocabulary			

SUMMARY TABLE OF PERCENTAGES

Level	Word Recognition	Oral Comprehension	Silent Comprehension	Average Comprehension
PP				
P				
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				

APPENDIX B

Reading Comprehension
Instructional Levels

Student Number	IRI (I)	CAT (C)	I-M	C-M
1	2.5	2.6	-1.96	-1.66
2	3.5	2.0	- .96	-2.26
3	5.5	1.6	+1.04	-2.66
4	4.5	2.7	+ .04	-1.56
5	6.5	7.2	+2.04	+2.94
6	3.5	4.1	- .96	- .16
7	6.5	9.2	+2.04	+4.94
8	4.5	3.7	+ .04	- .56
9	3.5	2.9	- .96	-1.36
10	4.5	2.2	+ .04	-2.06
11	3.5	5.2	- .96	+ .94
12	1.5	1.9	-2.96	-2.36
13	7.5	7.0	+3.04	+2.74
14	5.5	3.8	+1.04	- .46
15	5.5	11.1	+1.04	+6.84
16	7.5	7.3	+3.04	+3.04
17	5.5	6.7	+1.04	+2.44
18	4.5	4.1	+ .04	- .16
19	5.5	7.2	+1.04	+2.94
20	.5	2.1	-3.96	-2.16
21	4.5	2.2	+ .04	-2.06
22	3.5	2.1	- .96	-2.16
23	4.5	1.5	+ .04	-2.76
24	5.5	3.2	+1.04	-1.06
25	1.5	2.9	-2.96	-1.36
Total 25	111.5	106.5	.00	.00
Mean (M)	4.46	4.26		
Std. Dev.	1.77	2.67		
Corr.	.62			